

# **Inference of Biogeographical Ancestry and Pigmentation Phenotype using Single Nucleotide Polymorphisms**

**Charmain Vanessa Castel**

A thesis submitted in fulfilment of the requirements for the degree of  
Doctor of Philosophy



Centre for Forensic Science

University of Technology, Sydney, NSW, Australia

2014

## **Certificate of Original Authorship**

---

I certify that work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

**Charmain Vanessa Castel**

July 2014

## Acknowledgements

---

First and foremost, I would like to sincerely thank my supervisor Associate Professor Anita Piper for her ongoing support and invaluable guidance throughout my candidature. I have thoroughly enjoyed working with you over the years and have learnt so much, not only about the ‘science’ but also myself and my capabilities. Thank you for helping me to grow into an independent researcher and for your patience and continuous encouragement.

Thank you to Prof. Claude Roux (Centre for Forensic Science) and the Faculty of Science at the University of Technology, Sydney (UTS) for their funding and support of this project. I would also like to thank the UTS Graduate School who have awarded me with an Australian Postgraduate Award and have provided funding support to attend national and international conferences.

Thank you to Dr Runa Daniel (Victoria Police Forensic Services Centre) for the advice she provided in the early stages of this project, particularly for showing me the ins and outs of the ABI 310. Special thanks also extend to Connie Akratos and Zofia Winiarski from Technical Services (UTS) for kindly providing me with some of the materials and equipment I needed to complete the assay validation phase of this project. I would also like to thank all of the volunteers who provided DNA samples and without whom this project would not have been possible.

Thank you to all my friends and colleagues at UTS, past and present. In particular, Dr Michael Johnson, Dr Marilyn Katrib, Dr Rob Walker, Dr Michael Lees, Dr Nicholas Archer and Prue Gatt for convincing me to do a PhD. Whilst this has at times been a tough journey, I am so glad I took the bait and could not have done it without your encouragement, guidance and wisdom. Thank you all so much for the fun times and memories that I will cherish for a lifetime. Most importantly, I would like to thank Dr Stephanie Dowdell for being my office buddy, for all the teas and invaluable words of wisdom, for your kindness and generosity and for all the guidance, moral support and encouragement that you have given me over the years. I am so lucky to have made a lifetime friend like you!

Thank you to my dearest and oldest friend Mohini Saini for 15 years of friendship, laughter and fun. Despite the distance that now separates us, you have not ceased to provide me with your ongoing support and encouragement all the way from the other side of the World and I truly appreciate all that you have done.

I would especially like to thank (fellow soon to be Dr) Michael Strauss for sharing this incredible journey with me and for being by my side from day one. Thank you for being an amazing and understanding person, for your patience and for always supporting, and encouraging me through it all, particularly the difficult times. I would not have made it without you and I look forward to starting the next chapter of our lives together.

Finally, I would like to thank my family, especially my parents Yves Castel, Lou and Tony Doyle and my sister Karen for their wisdom and ongoing support and throughout my life. Each of you have inspired me in your own special way, but you have all taught me to never give up and that no challenge is too great. Your unwavering belief in me and my abilities astounds me and I would not have accomplished this, one of the greatest intellectual challenges of my life, without you all. I am truly grateful to have been blessed with a family like you.

## Table of Contents

---

<b>Certificate of Original Authorship .....</b>	<b>ii</b>
<b>Acknowledgements.....</b>	<b>iii</b>
<b>Table of Contents .....</b>	<b>v</b>
<b>Table of Figures.....</b>	<b>xii</b>
<b>Table of Tables .....</b>	<b>xvii</b>
<b>Publications .....</b>	<b>xix</b>
<b>Presentations.....</b>	<b>xix</b>
<b>List of Abbreviations .....</b>	<b>xx</b>
<b>Abstract .....</b>	<b>xxiii</b>
<b>Chapter 1. Introduction.....</b>	<b>2</b>
1.1 Preface.....	2
1.2 Tracing the origins and phylogeny of modern humans using haploid markers .....	3
1.2.1 MtDNA and Y chromosomal variation: .....	4
1.2.2 Uncovering the origin of modern humans using mtDNA and NRY SNPs .....	10
1.2.3 Reconstructing human mitochondrial and NRY phylogeny .....	14
1.2.4 Global distribution of mtDNA and NRY haplogroups .....	20
1.3 Autosomal SNPs .....	22
1.3.1 Autosomal SNPs: ancestry informative markers .....	22
1.3.2 Autosomal markers of human pigmentation .....	24
1.4 SNP Databases .....	24
1.5 Applications of genetic polymorphisms in forensic casework: .....	25
1.5.1 SNP analysis in forensic identification .....	26
1.5.2 Applications of genetic polymorphisms in forensic intelligence .....	28
1.5.2.1 Prediction of biogeographical ancestry and physical appearance .....	28
1.5.2.2 Genotyping strategies .....	32
1.5.2.3 Assignment methods .....	34

1.6	Project background.....	38
1.7	Objectives.....	40
<b>Chapter 2.</b>	<b>Materials and Methods.....</b>	<b>43</b>
2.1	Materials.....	43
2.1.1	Reagents, kits and consumables .....	43
2.1.2	Enzymes .....	44
2.1.3	Solutions.....	44
2.1.4	Synthetic Oligonucleotides .....	44
2.1.5	Population samples.....	50
2.1.6	Software .....	52
2.2	Methods.....	53
2.2.1	Multiplex SNP Assay design.....	53
2.2.1.1	Multiplex polymerase chain reaction (PCR) primer design .....	53
2.2.1.2	Multiplex single base extension (SBE) primer design .....	54
2.2.1.3	SNaPshot <sup>®</sup> Primer Focus reaction .....	55
2.2.2	Multiplex Assay optimisation and genotyping.....	56
2.2.2.1	Isolation and purification of buccal cells.....	56
2.2.2.2	Polymerase Chain Reaction (PCR) .....	57
2.2.2.3	Agarose gel electrophoresis.....	58
2.2.2.4	Post-PCR purification of DNA template.....	58
2.2.2.5	SNaPshot <sup>®</sup> reaction .....	58
2.2.2.6	Post- extension treatment .....	61
2.2.2.7	Capillary electrophoresis and Genescan <sup>®</sup> Analysis.....	61
2.2.2.8	Haplogroup and genotype determination .....	63
2.2.3	Statistical Analysis .....	63
2.2.3.1	Calculating haplogroup, allele and genotype frequencies.....	63
2.2.3.2	Assessing mtDNA and NRY population diversity.....	64

2.2.3.3	Assessing population diversity using autosomal SNPs .....	65
2.2.3.4	Analysis of Molecular Variance (AMOVA) .....	65
2.2.3.5	Genetic Distance and Number of Migrants .....	66
2.2.3.6	Principal Co-ordinates Analysis (PCoA).....	66
2.2.3.7	Neighbour-joining (NJ) trees.....	66
2.2.3.8	STRUCTURE analysis .....	67
2.2.3.9	Investigating departures from independence.....	69
2.2.3.10	Inference of bi-parental BGA using the Snipper App Suite 2.0.....	70
2.2.3.11	Inference of bi-parental BGA using GenAlEx v6.501 .....	70
2.2.3.12	Assessing allelic associations of nine pigmentation SNPs with external visible characteristics.....	71
2.2.3.13	Inference of pigmentation phenotype and external visible characteristics. ....	73
2.2.4	Sensitivity testing .....	74
2.2.5	Preliminary validation of the forensic intelligence 5-multiplex SNP genotyping system for forensic casework .....	74
2.2.5.1	Mock casework DNA samples .....	74
2.2.5.2	Genomic DNA extraction of casework DNA samples.....	75
2.2.5.3	Genomic DNA quantitation.....	76
2.2.5.4	Genotyping mock casework samples .....	77
<b>Chapter 3. Multiplex SNP Assay Development (Stage 1): Multiplex 1 to 4 for the Inference of Maternal and Paternal Ancestry .....</b>		<b>79</b>
3.1	Introduction.....	79
3.2	SNP Assay development- Stage 1 .....	80
3.2.1	Haploid SNP loci selection .....	80
3.2.2	Hierarchical multiplex assay design.....	87
3.2.3	Multiplex primer design .....	88
3.3	Results and Discussion.....	93

3.3.1	Multiplex 1 assay optimisation .....	93
3.3.1.1	Singleplex and multiplex amplification of Multiplex 1 PCR primers.....	93
3.3.1.2	Multiplex 1 PCR optimisation- Adjusting magnesium and dNTP concentration .....	96
3.3.1.3	Multiplex 1 PCR optimisation- Adjusting Taq polymerase concentration .. .....	97
3.3.1.4	Multiplex 1 PCR optimisation- Adjusting PCR buffer concentration: ....	99
3.3.1.5	Singleplex extension of Multiplex 1 SBE primers using Multiplex 1 PCR template .....	99
3.3.1.6	Multiplex 1 SBE reactions using Multiplex 1 PCR product as template. ... .....	100
3.3.1.7	Investigating the failed extension of the C10400T, C3970T, and T12705C alleles .....	103
3.3.1.8	Singleplex extension of modified multiplex 1 SBE primers using the modified multiplex 1 PCR product as template .....	107
3.3.1.9	Equalisation of allele signals through the optimisation of Multiplex 1 PCR and SBE primer concentrations .....	110
3.3.1.10	Effect of polymer concentration on the electrophoretic mobility of Multiplex 1 SBE primers.....	113
3.3.2	Multiplex 2 assay optimisation .....	116
3.3.2.1	Preliminary optimisation of the Multiplex 2 PCR reaction.....	116
3.3.2.2	Optimisation of the multiplex 2 SNaPshot™ reaction .....	118
3.3.3	Multiplex 3 assay optimisation .....	122
3.3.3.1	Preliminary optimisation of the Multiplex 3 PCR reaction.....	122
3.3.3.2	Optimisation of the Multiplex 3 SNaPshot™ reaction.....	125
3.3.4	Multiplex 4 assay optimisation .....	128
3.3.4.1	Preliminary optimisation of the Multiplex 4 PCR reaction.....	128
3.3.4.2	Optimisation of the multiplex 4 SNaPshot™ reaction .....	130
3.3.5	Initial population testing.....	133



3.4	Summary .....	138
<b>Chapter 4. Multiplex SNP Assay Development (Stage 2): Multiplex 5 for the Inference of Bi-parental Ancestry and Pigmentation Phenotype.....141</b>		
4.1	Introduction .....	141
4.2	SNP Assay development- Stage 2 .....	142
4.2.1	Autosomal SNP selection.....	142
4.2.1.1	AIMs in blood grouping and cytokine genes.....	144
4.2.1.2	SNPs in Human pigmentation genes .....	150
4.2.1.2.1	Fixed difference markers in the SLC24A5 and SLC45A2 genes ....	153
4.2.1.2.2	Polymorphisms in the MC1R gene .....	153
4.2.1.2.3	SNPs in the OCA2 gene.....	155
4.2.1.2.4	Polymorphisms in the HERC2 gene .....	156
4.2.1.2.5	Agouti Signalling Protein polymorphisms .....	158
4.2.2	Multiplex primer design .....	158
4.3	Results and Discussion.....	159
4.3.1	Multiplex 5 assay optimisation .....	159
4.3.1.1	Preliminary optimisation of the Multiplex 5 PCR reaction.....	159
4.3.1.2	Optimisation of the Multiplex 5 SNaPshot™ reaction.....	163
4.3.1.2.1	Singleplex extension of Multiplex 5 SBE primers using preliminary Multiplex 5 PCR product as template.....	163
4.3.1.2.2	Multiplex extension of Multiplex 5 SBE primers using preliminary Multiplex 5 PCR product as template.....	165
4.3.1.2.3	Investigation of the failed extension of the ASIP SBE primer .....	167
4.3.1.2.4	Redesigning the ASIP SBE primer .....	173
4.3.1.2.5	Singleplex extension of modified Multiplex 5 SBE primers using Multiplex 5 PCR product as template.....	175
4.3.1.2.6	Incorporation of the modified Multiplex 5 SBE primers into the Multiplex 5 SNaPshot™ reaction .....	177

4.3.1.2.7	Optimisation of 12-Plex SNaPshot™ reactions excluding ASIP ....	180
4.3.1.2.8	Optimisation of the ASIP singleplex SNaPshot™ reaction.....	183
4.3.1.2.9	Fine tuning of the Multiplex 5 ASIP singleplex and 12-Plex SNaPshot™ reactions .....	184
4.3.2	Initial population testing.....	185
4.3.3	Simultaneous separation and detection of the ASIP singleplex and multiplex 5 12-Plex SNaPshot™ reactions using capillary electrophoresis .....	188
4.4	Summary .....	192
<b>Chapter 5. Preliminary validation of the 5-multiplex SNP assay for forensic intelligence .....</b>		<b>195</b>
5.1	Introduction.....	195
5.2	Results.....	196
5.2.1	Large- scale genotyping of population samples.....	196
5.2.2	Sensitivity testing .....	197
5.2.3	Distribution of mtDNA, NRY and autosomal SNP alleles in test population groups .....	201
5.2.3.1	mtDNA and NRY haplogroups .....	201
5.2.3.2	Autosomal allele and genotype frequencies .....	203
5.2.4	Genetic Diversity in six major Australian sub-populations .....	212
5.2.4.1	MtDNA and NRY diversity.....	212
5.2.4.2	Autosomal diversity.....	214
5.2.5	Analysis of population structure using mtDNA and NRY SNPs.....	216
5.2.6	Analysis of population structure using autosomal SNPs .....	223
5.2.7	Assessing departures from independence at autosomal SNP loci.....	231
5.2.8	Inference of maternal and paternal BGA .....	233
5.2.9	Inference of bi-parental BGA using Multiplex 5 .....	237
5.2.9.1	Inference of bi-parental BGA using the Snipper App Suite 2.0.....	237
5.2.9.2	Inference of bi-parental BGA using GenAlEx v6.5 .....	239

5.2.10 Predictive power of the 5-multiplex SNP genotyping system for the inference of regional biogeographical ancestry .....	241
5.2.11 Inference of pigmentation phenotype and EVCs using nine SNPs in pigmentation genes.....	246
5.2.12 Genotyping mock forensic casework samples .....	252
5.3 Summary .....	260
<b>Chapter 6. General Discussion.....</b>	<b>263</b>
6.1 Overview of the project.....	263
6.2 Performance of the 5-multiplex SNP assay during population testing .....	264
6.3 Population genetic analyses of mtDNA, NRY and autosomal haplotype and genotype data .....	267
6.4 Inference of biogeographical ancestry using the 5-multiplex SNP genotyping system.....	277
6.5 Inference of external visible characteristics using nine pigmentation associated SNPs.....	284
6.6 Assay sensitivity and application in forensic casework.....	289
6.7 Conclusions and Future Directions .....	292
<b>References .....</b>	<b>295</b>

## Table of Figures

---

Figure 1.1: Model of autosomal, Y-chromosomal and mitochondrial DNA inheritance.	5
Figure 1.2: Schematic of the human Y chromosome, depicting major regions and active genes.....	6
Figure 1.3: Schematic of the human mitochondrial DNA genome, showing major regions and active genes.....	8
Figure 1.4: The migration of modern humans .....	14
Figure 1.5: Simplified schematic of the mtDNA phylogenetic tree.....	17
Figure 1.6: Simplified schematic of the NRY phylogenetic tree. ....	19
Figure 2.1: Schematic of the SNaPshot <sup>®</sup> reaction.....	59
Figure 3.1: Phylogenetic tree of the 21 selected mtDNA SNPs showing the geographical affiliation of the defining haplogroups.....	85
Figure 3.2: Phylogenetic tree of the 28 selected NRY SNPs showing the geographical affiliation of the defining haplogroups.....	86
Figure 3.3: MtDNA and NRY SNP phylogenies and haplogroup defining haplotypes for Multiplex 1.....	89
Figure 3.4: MtDNA and NRY SNP phylogenies and haplogroup defining haplotypes for Multiplex 2.....	90
Figure 3.5: MtDNA and NRY SNP phylogenies and haplogroup defining haplotypes for Multiplex 3.....	91
Figure 3.6: MtDNA and NRY SNP phylogenies and haplogroup defining haplotypes for Multiplex 4.....	92
Figure 3.7: Multiplex 1 singleplex and multiplex PCR products.....	95
Figure 3.8: Multiplex 1 PCR optimisation of magnesium and dNTP concentration. ....	97
Figure 3.9: Multiplex 1 PCR optimisation of <i>Taq</i> polymerase concentration. ....	98
Figure 3.10: Multiplex 1 PCR optimisation of PCR buffer concentration. ....	98
Figure 3.11: Capillary electropherograms of singleplex extensions of 0.2µM Multiplex 1 SBE primers using Multiplex 1 PCR product as template.....	101
Figure 3.12: Capillary electropherograms of singleplex extensions of 2µM Multiplex 1 SBE primers using Multiplex 1 PCR product as template.....	102
Figure 3.13: Capillary electropherograms of Multiplex 1 SBE primer optimisations using Multiplex 1 PCR product as template. ....	104
Figure 3.14: Autodimer result for C3970T PCR forward primer dimer screen.....	106

Figure 3.15: Amplification of the original and modified C3970T PCR primer pair in duplex and Multiplex 1 PCR reactions. ....	106
Figure 3.16: Capillary electropherograms of the singleplex extension of 0.2μM C3970T SBE primer using singleplex PCR product as template. ....	108
Figure 3.17: Autodimer result for C3970T SBE primer hairpin screen. ....	108
Figure 3.18: Capillary electropherograms of singleplex extensions of 0.2μM of the modified C3970T, T12705C and C10400T Multiplex 1 SBE primers using Multiplex 1 PCR product as template. ....	109
Figure 3.19: Capillary electropherograms for the optimisation of Multiplex 1 12-Plex SNaPshot™ reactions incorporating the redesigned SBE primers and using Multiplex 1 PCR product as template. ....	111
Figure 3.20: Capillary electropherograms of multiplex 1 SNaPshot™ reactions separated in (A) POP-4™ and (B) POP-6™ polymer. ....	115
Figure 3.21: Multiplex 2 singleplex and multiplex PCR products. ....	117
Figure 3.22: Optimisation of Multiplex 2 PCR primer concentrations. ....	118
Figure 3.23: Capillary electropherograms of singleplex extensions of 0.2μM multiplex 2 SBE primers using optimised Multiplex 2 PCR product as template. ....	119
Figure 3.24: Capillary electropherograms for the optimisation of Multiplex 2 SNaPshot™ reactions using Multiplex 2 PCR product. ....	121
Figure 3.25: Multiplex 3 singleplex and multiplex PCR products. ....	123
Figure 3.26: Optimisation of Multiplex 3 PCR primer concentrations. ....	124
Figure 3.27: Electropherograms of the singleplex amplification of 0.2μM multiplex 3 SBE primers using optimised Multiplex 3 PCR product as template. ....	126
Figure 3.28: Capillary electropherograms for the optimisation of Multiplex 3 SNaPshot™ reactions using Multiplex 3 PCR product as template. ....	127
Figure 3.29: Multiplex 4 singleplex and multiplex PCR products. ....	129
Figure 3.30: Optimisation of Multiplex 4 PCR primer concentrations. ....	130
Figure 3.31: Capillary electropherograms of singleplex extensions of 0.2μM Multiplex 4 SBE primers using optimised Multiplex 4 PCR product as template. ....	132
Figure 3.32: Capillary electropherograms for the optimisation of Multiplex 4 SNaPshot™ reactions using Multiplex 4 PCR product as template. ....	133
Figure 3.33: Initial population testing electropherograms of the Multiplex 1 and 2 profiles for five individuals with differing BGA. ....	136

Figure 3.34: Initial population testing electropherograms of the Multiplex 3 and 4 profiles for five individuals with differing BGA. ....	137
Figure 4.1: Overview of melanogenesis .....	151
Figure 4.2: Multiplex 5 singleplex and multiplex PCR products.....	160
Figure 4.3: Preliminary optimisation of Multiplex 5 PCR primer concentrations.....	162
Figure 4.4: Capillary electropherograms of the singleplex extension of 0.2µM Multiplex 5 SBE primers using Multiplex 5 PCR product as template.....	164
Figure 4.5: Capillary electropherograms for the optimisation of Multiplex 5 SNaPshot™ reactions using Multiplex 5 PCR product as template.....	166
Figure 4.6: Amplification of ASIP PCR primers in singleplex, 4-PlexA and 13-Plex PCR reactions at variable PCR primer concentrations. ....	169
Figure 4.7: Capillary electropherograms of the extension of the ASIP SBE primer in singleplex, 4-PlexA and 12-Plex SNaPshot™ reactions using singleplex and multiplex PCR products as template. ....	170
Figure 4.8: Optimisation of singleplex ASIP SNaPshot™ reactions using modifications to the SNaPshot™ cycling protocol and the addition of PCR adjuvants.....	172
Figure 4.9: Autodimer primer screen results for (A) the unmodified ASIP SBE primer and (B) the modified ASIP SBE primer with inserted mismatch adapted from Brudnik <i>et al.</i> 2009.....	174
Figure 4.10: Capillary electropherograms of the singleplex amplification of the modified Multiplex 5 SBE primers using Multiplex 5 PCR product as template.....	176
Figure 4.11: Capillary electropherograms for the optimisation of the Multiplex 5 13-Plex SNaPshot™ reaction incorporating the modified SBE primers and using Multiplex 5 PCR product as template.....	178
Figure 4.12: Capillary electropherograms for the separate optimisation of ASIP singleplex and Multiplex 5 12-Plex SNaPshot™ reactions using Multiplex 5 PCR product as template. ....	182
Figure 4.13: Initial population testing electropherograms of the Multiplex 5 profiles for five individuals with differing BGA. ....	187
Figure 4.14: Electropherograms of the final ASIP singleplex and Multiplex 5 12-Plex SNaPshot™ reactions in separate and combined capillary electrophoresis runs using variable injection times. ....	189
Figure 5.1: Multiplex 1 and 2 sensitivity profiles for a European individual genotyped using various amounts of starting DNA template.....	198

Figure 5.2: Multiplex 3 and 4 sensitivity profiles for a European individual genotyped using various amounts of starting DNA template.....	199
Figure 5.3: Multiplex 5 sensitivity profiles for a European individual genotyped using various amounts of starting DNA template.....	200
Figure 5.4: Geographic distribution of observed (A) mtDNA and (B) NRY haplogroups. ....	202
Figure 5.5: Allele and genotype frequencies for the MC1R (rs885479), DARC (rs2814778) and OCA2-1 (rs7495174) loci. ....	205
Figure 5.6: Allele and genotype frequencies for the HERC2 (rs12913832), MATP (rs16891982) and SLC24A5 (rs1426654) loci.....	207
Figure 5.7: Allele and genotype frequencies for the F13B (rs6003), OCA2-2 (rs4778241) and FUT2 (rs1800021) loci. ....	209
Figure 5.8: Allele and genotype frequencies for the OCA2-3 (rs4778138), OCA2-4 (rs1545397) and IL-6 (rs1800795) loci.....	210
Figure 5.9: Allele and genotype frequencies for the ASIP (rs6058017) locus. ....	212
Figure 5.10: AMOVA results for the analysis of six Australian sub-populations using 21 mtDNA and 28 NRY SNPs.....	217
Figure 5.11: Principal Coordinates Analysis of mtDNA and NRY genetic distances for six Australian sub-populations.....	220
Figure 5.12: Neighbour-joining tree for six Australian sub-populations based on calculated mtDNA and NRY genetic distances. ....	222
Figure 5.13: AMOVA results for the analysis of six Australian sub-populations using 13 autosomal SNPs. ....	224
Figure 5.14: Principal Coordinates Analysis of autosomal genetic distances for six Australian sub-populations. ....	226
Figure 5.15: Neighbour-joining tree for six Australian sub-populations based on calculated autosomal genetic distances.....	226
Figure 5.16: Estimation of the number of $K$ clusters in the autosomal dataset using the Evanno method.....	228
Figure 5.17: Estimation of population sub-structure within the autosomal dataset using STRUCTURE v2.3.4. ....	229
Figure 5.18: Individual genetic ancestry at $K=3$ in the 1000 Genomes (Phase I) dataset estimated by STRUCTURE v2.3.4 from 13 autosomal SNPs.....	238

Figure 5.19: GenALEx Pairwise population assignment plots for four major continental population groups.....	242
Figure 5.20: Preliminary validation of the 5-multiplex SNP genotyping system on forensic casework samples from Donor 1.....	254
Figure 5.21: Preliminary validation of the 5-multiplex SNP genotyping system on forensic casework samples from Donor 2.....	255
Figure 5.22: Preliminary validation of the 5-multiplex SNP genotyping system on forensic casework samples from Donor 3.....	256
Figure 5.23: Preliminary validation of the 5-multiplex SNP genotyping system on forensic casework samples from Donor 4.....	257
Figure 6.1: Updated Phylogenetic tree of the 21 selected mtDNA SNPs showing the geographical origin of the defining haplogroups. ....	266



## Table of Tables

---

Table 2.1: Designed (a) mtDNA, (b) NRY and (c) autosomal PCR primers.....	45
Table 2.2: Designed (a) mtDNA and NRY and (b) autosomal SBE primers .....	48
Table 2.3: Distribution of the Australian population samples genotyped using the developed 5-multiplex SNP genotyping system .....	51
Table 2.4: SNaPshot® Primer Focus™ reaction preparation .....	56
Table 2.5: Dye labels assigned to individual ddNTPs in the SNaPshot® reaction.....	59
Table 2.6: Preparation of SNaPshot® reaction controls .....	60
Table 2.7: Run parameters encoded by module GS POP-4 (1mL) E5.md5 used for electrophoresis of SNaPshot® products on the ABI PRISM® 310 Genetic Analyzer...	62
Table 2.8: Phenotype categories used in Fisher's Exact tests of allelic association on 40 Europeans. ....	72
Table 2.9 Hair, eye and skin colour categories used in population assignment tests for the inference of the pigmentation phenotype of 146 individuals using nine pigmentation SNPs.....	73
Table 2.10: Concentration of gDNA extracted from mock casework biological samples determined using the Qubit™ dsDNA HS Assay. ....	77
Table 3.1: Global distribution (%) of candidate mtDNA haplogroups.....	81
Table 3.2: Global distribution (%) of candidate NRY haplogroups. ....	82
Table 3.3: Comparison of expected and observed sizes of multiplex 1 SBE primers in POP-4™ and POP-6™ polymer. ....	115
Table 3.4: Final optimised PCR and SBE primer concentrations for Multiplex 1- 4 ...	135
Table 4.1: Multiplex 5 candidate autosomal SNPs .....	145
Table 4.2: Allele frequencies for Multiplex 5 candidate autosomal SNPs retrieved from the 1000 Genomes Database (Phase I) using SPSmart v5.1.1. ....	146
Table 4.3: Final optimised PCR and SBE primer concentrations for Multiplex 5 .....	185
Table 5.1: Diversity indices deduced from the analysis of (A) 21 mtDNA and (B) 28 NRY SNPs in 116 males from six populations.....	213
Table 5.2: Analysis of heterozygosity at 13 autosomal SNP loci in six major Australian sub-populations .....	215
Table 5.3: Genetic distance matrices of population pairwise $\Phi_{PT}$ and M values from the analysis of 21 mtDNA and 28 NRY SNPs in six Australian sub-populations.....	218

Table 5.4: Genetic distance matrix of population pairwise $F_{ST}$ and $M$ values from the analysis of 13 autosomal SNPs in six Australian sub-populations. ....	224
Table 5.5: Hardy-Weinberg equilibrium analysis of 13 autosomal SNP loci in six major Australian sub-populations .....	232
Table 5.6: Linkage equilibrium analysis of 13 autosomal SNP loci in six major Australian sub-populations .....	232
Table 5.7: Prediction correlation of the Multiplex 1 to 4 assays for the inference of maternal and paternal regional BGA.....	234
Table 5.8: Overall prediction correlation of Multiplex 1 to 4 for the inference of regional BGA .....	234
Table 5.9: Derived mtDNA haplogroups and prediction correlation for a sample of 30 females typed using Multiplex 1 to 4.....	236
Table 5.10: Results for the inference of bi-parental BGA using SNIPPER.....	238
Table 5.11: Results for the inference of bi-parental BGA using GenAlEx v6.5. ....	240
Table 5.12: Summary of the predictive power of the complete multiplex SNP assay (Multiplex 1 to 5) for the inference of regional BGA.....	243
Table 5.13: Assessing the predictive power of the 5-multiplex SNP genotyping system for the inference of BGA of female individuals.....	246
Table 5.14: Association of nine pigmentation SNPs with hair, eye and skin colour variation in Europeans (N=40).....	248
Table 5.15: Inference of EVCs using only pigmentation SNPs demonstrating positive associations to hair, eye and skin colour variation.....	250
Table 5.16: Inference of EVCs using nine pigmentation SNPs.....	252
Table 5.17: Inferred regional BGA of simulated casework sample donors.....	259
Table 5.18: Inferred hair, eye and skin colour of simulated casework sample donors.	259

## **Publications**

---

**Castel C.V.** and Piper A.A (2011) Development of a SNP multiplex assay for the inference of biogeographical ancestry and pigmentation phenotype. *Forensic Sci Int Genet Suppl Series*, **3**, e411-412

## **Presentations**

---

**Castel, C.V.** and Piper, A.A. Development of a SNP multiplex assay for the inference of biogeographical ancestry and pigmentation phenotype. Poster presented at the 24<sup>th</sup> World Congress of the International Society of Forensic Genetics, Vienna, Austria, August 2011. Abstract P014.

**Castel, C.V.** and Piper, A.A. Mitochondrial single nucleotide polymorphisms: Informative markers of biogeographical ancestry. Poster presented at the 20<sup>th</sup> International Symposium on the Forensic Sciences: Forensic Science on trial, Sydney, Australia, September 2010. Abstract 238.

## List of Abbreviations

---

AIM	Ancestry Informative Marker
$\alpha$ -MSH	$\alpha$ -Melanocyte Stimulating Factor
ASCCEG	Australian Standard Classification of Cultural and Ethnic Groups
ASIP	Agouti Signalling Protein
ATPase	ATP Synthase
AUC	Area Under the receiver characteristic operating Curve
BGA	Biogeographical Ancestry
bp	Base Pair
cAMP	Cyclic Adenosine Monophosphate
CEPH	Centre d'Etude du Polymorphisme Humain
CI	Confidence Interval
CIP	Calf Intestinal Phosphatase
CO	Cytochrome C oxidase
Cytb	Cytochrome b
DARC	Duffy Antigen/Receptor for Chemokine
dbSNP	SNP database
DCT	Dopachrome Tautomerase
$\delta$	Absolute Allele Frequency Difference
ddNTP	Dideoxyribonucleotide Triphosphate
dH <sub>2</sub> O	Deionised water
DHPLC	Denaturing High Performance Liquid Chromatography
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
EVC	External Visible Characteristic
<i>ExoI</i>	Exonuclease I
FIC	Fisher Information Content
[F]ddNTP	Fluorescently labelled Dideoxyribonucleotide Triphosphate
F <sub>ST</sub>	F-Statistic

FUT	Fucosyltransferase
GDA	Genetic Distance Analysis
gDNA	Genomic Deoxyribonucleic Acid
H <sub>e</sub>	Expected Heterozygosity
HGDP	Human Genome Diversity Project
HLTF	Helicase-Like Transcription Factor
H <sub>o</sub>	Observed Heterozygosity
HWE	Hardy Weinberg Equilibrium
I <sub>n</sub>	Rosenberg's Informativeness for Assignment measure
IL	Interleukin
LD	Linkage Disequilibrium
LE	Linkage Equilibrium
LEF1	Lymphoid Enhancer Binding Factor 1
MA	Minor Allele
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation-Time Of Flight
MATP	Membrane Associate Transporter Protein
MC1R	Melanocortin-1 Receptor
MCMC	Markov Chain Monte Carlo
MgCl <sub>2</sub>	Magnesium Chloride
MITF	Microphthalmia Transcription Factor
MLE	Maximum Likelihood Estimate
mM	Millimolar
mtDNA	Mitochondrial Deoxyribonucleic Acid
NCBI	National Center for Biotechnology Information
ND	NADH dehydrogenase
nDNA	Nuclear Deoxyribonucleic Acid
ng	Nanogram
NJ	Neighbour-Joining
NRY	Non-Recombining region of the Y-Chromosome
nt	Nucleotide
OCA2	Oculocutaneous Albinism Type 2
OR	Odds Ratio

PAR	Pseudoautosomal Region
PCoA	Principal Co-ordinates Analysis
PCR	Polymerase Chain Reaction
pg	Picogram
RFLP	Restriction Fragment Length Polymorphism
RFU	Relative Fluorescence Unit
rRNA	Ribosomal Ribonucleic Acid
SAP	Shrimp Alkaline Phosphatase
SBE	Single Base Extension
SIC	Shannon Information Content
SLC24A5	Solute Carrier Family 24 member 5
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
TBE	Tris Borate EDTA
TE	Tris EDTA
T <sub>m</sub>	Theoretical Melting Temperature
TMRCA	Time of the Most Recent Common Ancestor
TYR	Tyrosine
TYRP1	Tyrosine Related Protein-1
µg	Microgram
µL	Microlitre
USA	United States of America
UTS	University of Technology, Sydney

## Abstract

---

Conventional DNA profiling of Short Tandem Repeats (STR) provides little evidentiary value in the absence of reference profiles or in the case of a non-match. Recently, the forensic DNA intelligence field has flourished to provide investigators with valuable information from DNA samples that can narrow the collection of potential matches by identifying previously unknown reference individuals. Intelligence data of special interest includes the biogeographical ancestry (BGA) and external visible characteristics (EVC) such as the eye, hair and skin colour of unknown DNA samples donors.

Innovative technological advances like next-generation sequencing and microarrays have been crucial to the establishment of population diversity repositories comprising millions of DNA markers, the most abundant of which include single nucleotide polymorphisms (SNPs). Large-scale SNP studies of global populations have enabled reconstructions of mitochondrial (mtDNA) and non-recombining Y-chromosome (NRY) phylogenies, providing highly comprehensive population specific patterns of maternal and paternal genetic variation. Similarly, numerous patterns of autosomal genetic variation have been identified between different populations. These studies have culminated in panels of markers capable of resolving ancestry at the continental level. The identification of autosomal SNPs associated with human pigmentation variation has also resulted in the discovery of specific SNPs capable of predicting EVCs. Several DNA intelligence and phenotyping assays for the inference of BGA and for the prediction of eye, hair and skin colour have subsequently been developed. However, most of these intelligence tools have primarily focused on the analysis of one class of SNPs, hence limiting the amount of ancestry intelligence that could be obtained. The scarcity and often environmentally compromised nature of forensic biological evidence means that performing numerous individual intelligence tests is not optimal and a consolidated DNA intelligence diagnostic test is very much needed.

This study aimed to develop a SNP genotyping system that combined autosomal, NRY and mtDNA markers for comprehensive predictions of BGA and EVCs. Candidate SNPs were selected through literature and database searches to identify loci exhibiting skewed allele frequency differences between Sub-Saharan African, North African, Middle Eastern, European, South and East Asian populations. A hierarchical

arrangement comprising five separate multiplexes was implemented, in which SNP typing was performed by single-base extension assays. The haploid mtDNA and NRY SNPs were grouped into Multiplex 1 to 4, with SNPs defining maternal and paternal lineages (haplogroups) affiliated with the same geographic region grouped in the same reaction. The markers defining basal haplogroups were included in Multiplex 1, which is then used to identify the subsequent multiplex(es) required to achieve further haplogroup resolution and to minimise the number of tests required. The autosomal SNPs are typed separately in Multiplex 5.

A performance evaluation of the 5-multiplex SNP assay was undertaken on 146 individuals originating from the six major population groups of interest. Population genetic analyses of the mtDNA and NRY haplotypes and autosomal genotypes revealed that a greater degree of population differentiation was achieved with the selected NRY and autosomal SNPs than with the mtDNA SNPs. Moreover, the results indicated that the assay primarily allowed for the differentiation of continental ancestry, with populations in close proximity within continents, such as Europe, the Middle East and South Asia, often difficult to distinguish. However, the observed correlation between the declared and inferred geographic regions of maternal and paternal origin was high; 73-100% for maternal and 79-100% for paternal regional BGA. The bi-parental BGA predictions ranged from 85 to 95%, provided Middle Easterners and Europeans were grouped into a single Western Eurasian population. In 99% of cases, two of the three SNP classes correctly predicted the same ancestry from one of the five broad geographical regions (Sub-Saharan Africa, North Africa, Western Eurasia, South Asia and East Asia). High prediction accuracies were also observed for the inference of EVCs including hair (86-88%) and eye colour (81-95%). The DNA intelligence assay also demonstrated advanced performance with low starting amounts of genomic DNA, with full profiles observed for up to 100pg of template and for the analysis of routine casework biological samples. Consequently, this study presented the successful development of a novel, consolidated DNA intelligence tool that has displayed high performance for the inference of regional (continental) BGA and EVC in preliminary tests. Further validations of the assay are required; however the developed 5-multiplex SNP assay remains a valuable DNA intelligence diagnostic tool for the forensic science community.